IS 6733
Deep Learning on Cloud Platforms

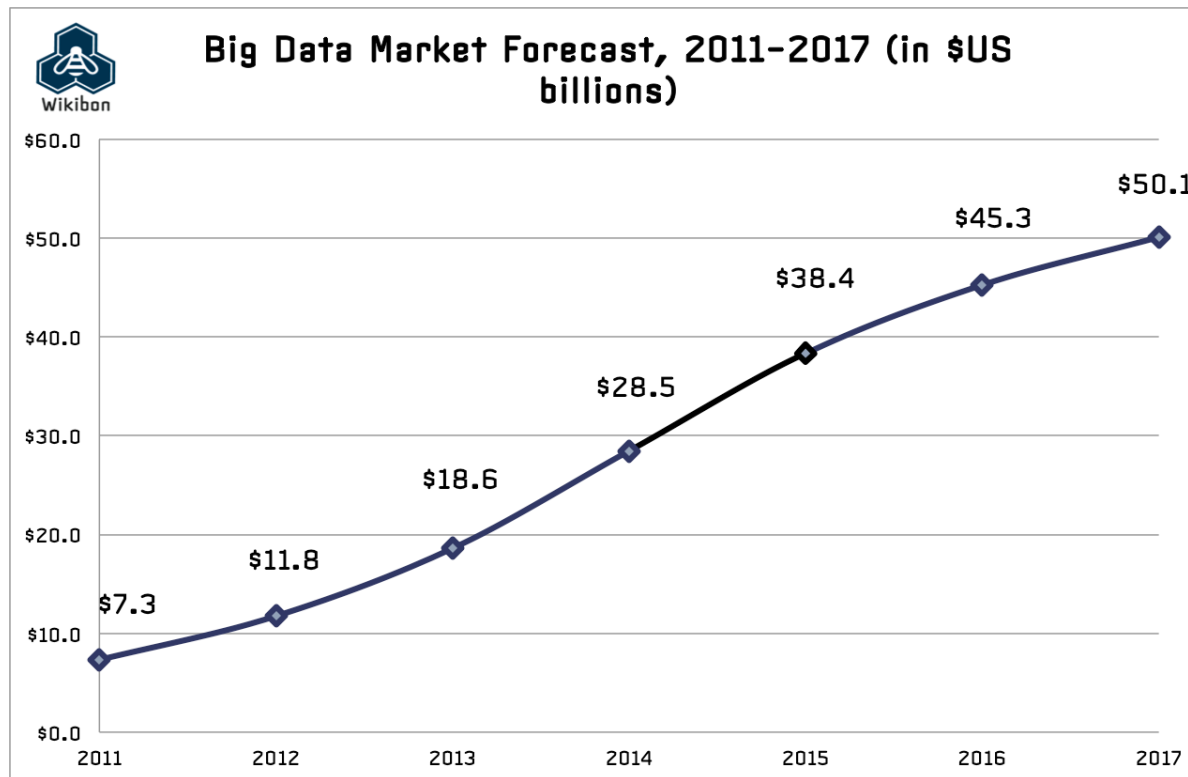## Lecture 1b

Big Data Characteristics and Challenges

Dr. Yuanxiong Guo

Department of Information Systems and Cyber Security

Big Data Market

Analysts forecast continued growth in the big data market

Big Data Market Forecast, 2011-2017 (in $US billions)

Who is Generating Big Data?

Social media and networks
(all of us are generating data)

Short Response at Large Scale

Mobile devices
(tracking all objects all the time)

Sensor technology and networks
(measuring all kinds of data)

Scientific instruments
(collecting all sorts of data)

Big Data: By the Numbers

- Google process 24PB of data per day (1 PB = 10^3 TB)
- AT&T transfers 30PB of data per day
- 90 trillion emails are sent per year
- Square Kilometer Array Telescope will be 100 TB/second
- Every second, on average, around 6,000 tweets are tweeted on Twitter (200 billion tweets per year)
- 300 hours of videos are uploaded to YouTube every minute; 5 billion videos are watched on Youtube every single day

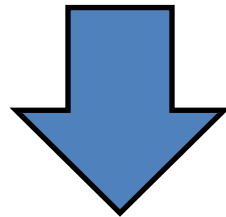# With Big Data, We've Moved into a New Era of Analytics

- Key enablers for the appearance and growth of big data are
  - Increase in storage capabilities and processing power
  - Availability of data
  - Advanced algorithms
  - Protecting data (security and privacy)

Characteristics of Big Data

- **Volume**: vast amounts of data that is generated
- **Velocity**: the speed at which data is being generated and moves from one point to the next
- **Variety**: the ever increasing different data formats (i.e., text, video, image, voice, geospatial)
- **Veracity**: unmeasurable uncertainties and truthfulness and trustworthiness of data
- **Valence**: the connectedness of big data in the form of graphs

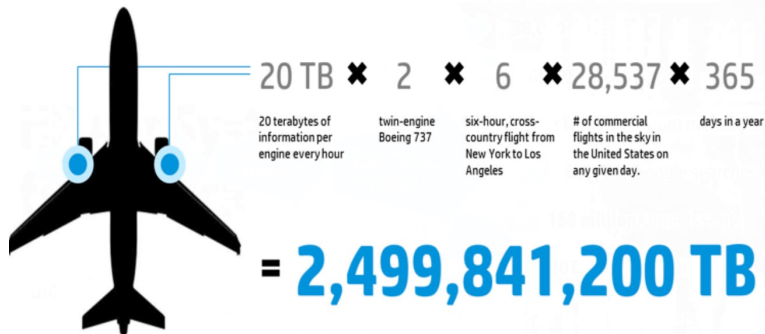**Value**

Volume = Size & Scale

Examples

Every minute ..

- **204 million emails on Gmail**
- **200,000 Facebook photos**
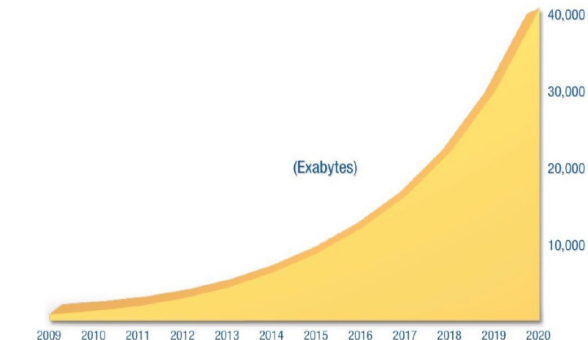- **1.3 million video views on YouTube**

Exponential Data Growth

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-...

Business Inside

Sensor data from a cross-country flight

20 TB ✖ 2 ✖ 6 ✖ 28,537 ✖ 365

| 20 terabytes of information per engine every hour | twin-engine Boeing 737 | six-hour, cross-country flight from New York to Los Angeles | # of commercial flights in the sky in the United States on any given day. | days in a year |

= 2,499,841,200 TB

Challenges

Storage Capacity Leads to

- **Processing**
- **Networking**
- **Access & Retrieval**
- **Performance**
- **Cost**
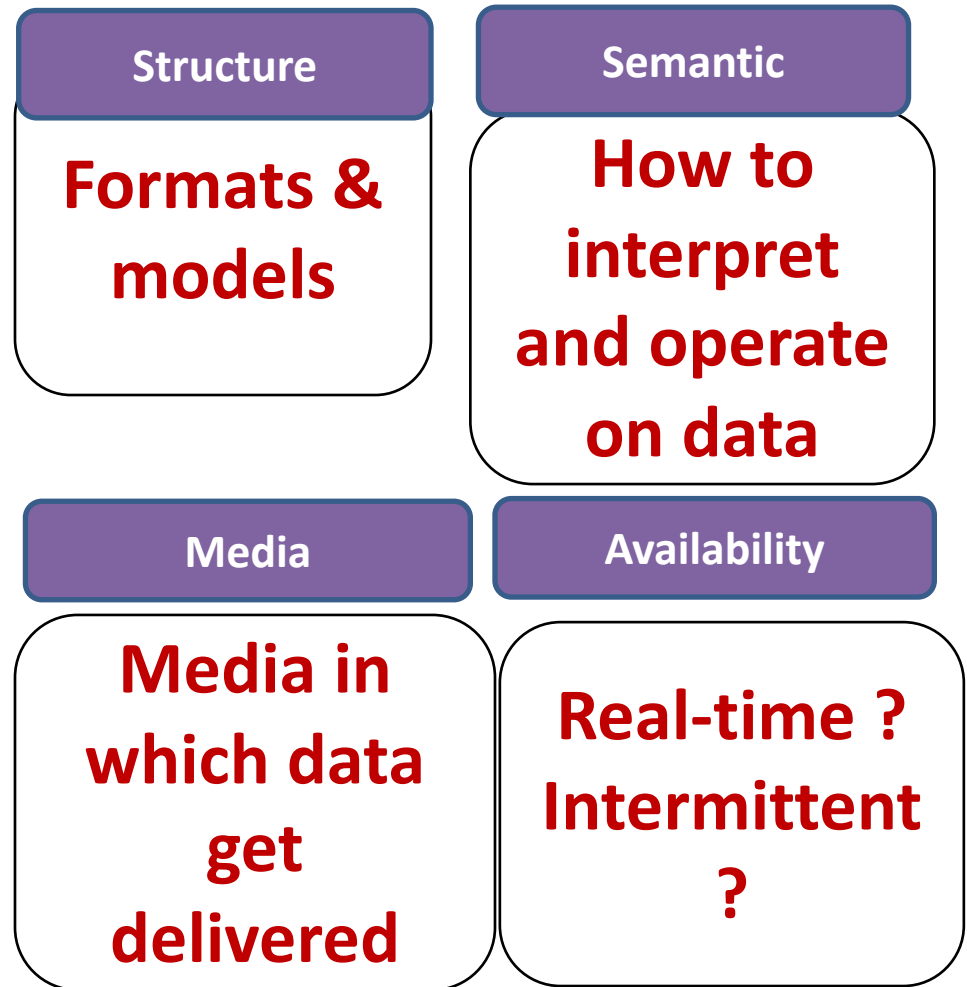
How Much Data are We Talking About?

- A DVD ~ 5 GBs

- 1 TB = 300 hours of good quality video

- IDC (International Data Corporation) says Digital Universe will be 35 Zettabytes by 2020
  - 1 ZB = 1 Billion TB

Variety = Heterogeneous Data Type -> Complexity

- Variety Issues
  - Difficult to create common storage
  - Management and policy challenges
  - Harder to process (compare and match data) and correlate

Structure

Formats & models

Semantic

How to interpret and operate on data

Media

Media in which data get delivered

Availability

Real-time ? Intermittent ?

Email: Variety within a Type

☆ **Wendy Frost** 📁 Inbox – Exchange 10:33 AM WF

Spread the Word! Fulbright U.S. Scholar Opportunities in France

As a Fulbright U.S. Scholar alum to France, I am writing to seek your assistance in spreading the word about the awards that we have available this year!

We are currently seeking US citizens who are interested in teaching or conducting research during the 2020-2021 academic year through the Fulbright U.S. Scholar Program. Our current application deadline is **September 16, 2019**. I wanted to highlight the awards that are available in France and for those interested in perusing wide range of offerings may do so in our awards catalog.

Award Offerings in France
All Disciplines
French Studies (Arts, Humanities and Social Sciences)
Fulbright-EHESS Post-Doctoral Award in Social Sciences (Junior Research)
Fulbright-IMéRA-Aix-Marseilles Université Chair in Migration Studies
Fulbright-Paris Seine University Award
Fulbright-Tocqueville Distinguished Chair (All Disciplines)
Grand Est Regional Award
Hauts-de-France Regional Award
Hauts-de-France Regional Award (Junior Research)
Nouvelle-Aquitaine Regional Award

Award Opportunity for International Education Administrators
U.S. France International Education Administrators Program (deadline for this award only = Monday, February 3, 2020)

For more information on what the application entails, I would recommend reading the application guidelines and listening to our upcoming/past webinars. I am happy to schedule a call to talk to you further about these opportunities.

Please do not hesitate to reach out and let me know if you or your colleagues have any questions! Thank you for encouraging your colleagues to apply!

Sincerely,
Cecilia

Cecilia P. Kocinski-Mulder
Program Officer, Europe/Eurasia
Fulbright Student and Scholar Programs
Institute of International Education (IIE)
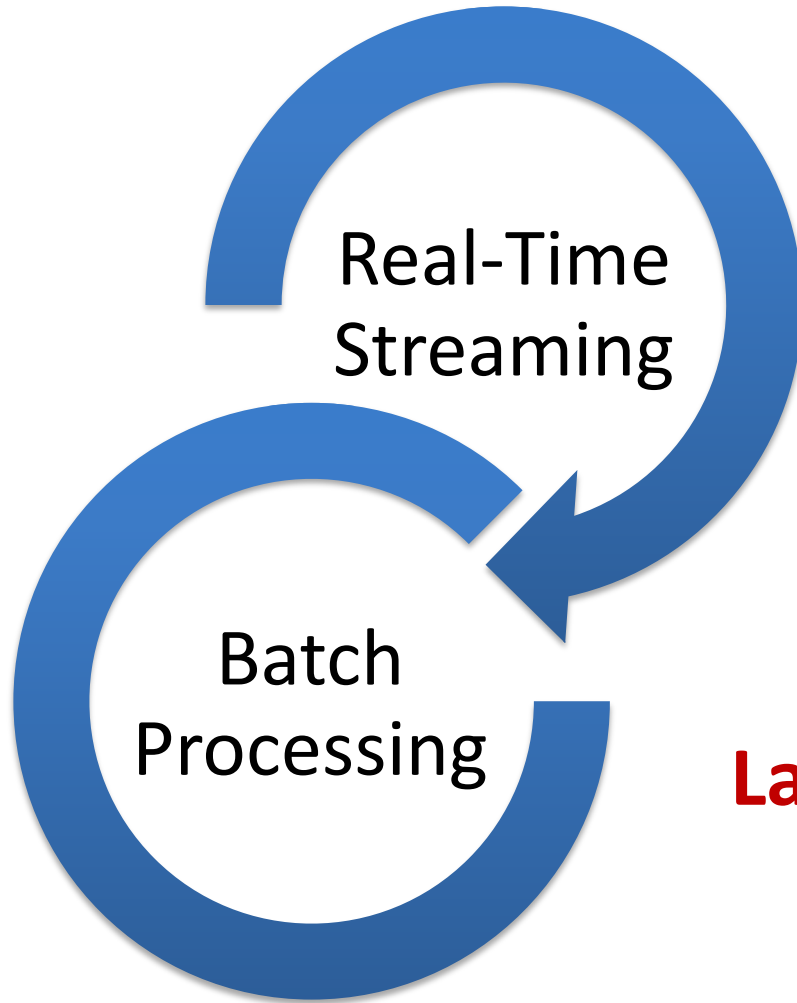1400 K Street NW, Ste 700
Washington, DC 20005
202-686-6240
ckocinskimulder@iie.org
IIE · *The Power of International Education*

Velocity

Real-Time Streaming

Batch Processing

Late decisions can lead to missing opportunity

Batch Processing

1) Collect Data
2) Clean Data
3) Process Data
4) Output Results
5) Action
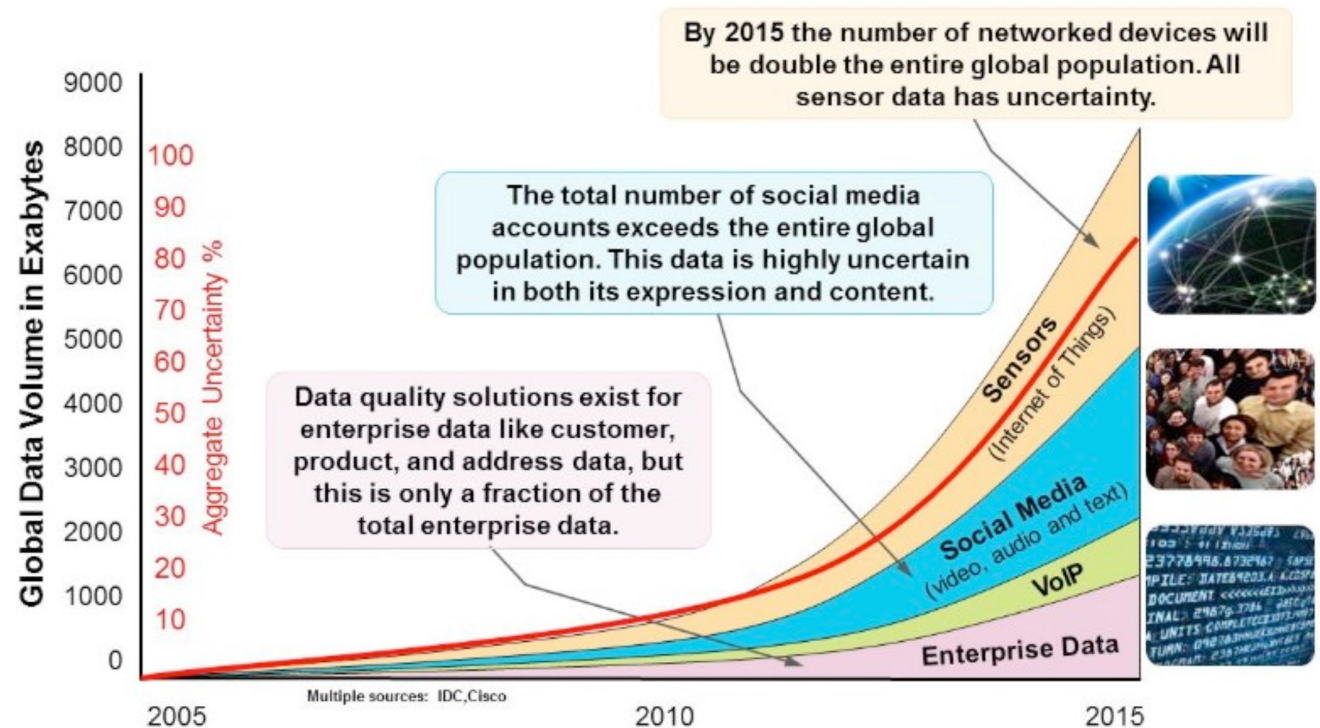
Real-Time Processing

1) Instantly capture streaming data
2) Filter real-time data for processing
3) Process real-time data
4) Action

Veracity = Quality (Validity & Volatility)

- Accuracy of data

- Reliability of the data source (Data Provenance)

- Context within Analysis

By 2015 the number of networked devices will be double the entire global population. All sensor data has uncertainty.

The total number of social media accounts exceeds the entire global population. This data is highly uncertain in both its expression and content.

Data quality solutions exist for enterprise data like customer, product, and address data, but this is only a fraction of the total enterprise data.

Global Data Volume in Exabytes

Aggregate Uncertainty %

Sensors (Internet of Things)

Social Media (video, audio and text)

VoIP

Enterprise Data

Multiple sources: IDC, Cisco

Hutzler 571 Banana Slicer
$3.99 ✓Prime | FREE One-Day

Delivered tomorrow for FREE with qualifying orders over $35.   Details | In Stock. Ships from and sold by Amazon.com. Gift-wrap available.

🛒 Add to Cart

⭐⭐⭐⭐⭐ **No more winning for you, Mr. Banana!**
By SW3K on March 3, 2011

Size: 10" | Item Package Quantity: 1

For decades I have been trying to come up with an ideal way to slice a banana. "Use a knife!" they say. Well...my parole officer won't allow me to be around knives. "Shoot it with a gun!" Background check...HELLO! I had to resort to carefully attempt to slice those bananas with my bare hands. 99.9% of the time, I would get so frustrated that I just ended up squishing the fruit in my hands and throwing it against the wall in anger. Then, after a fit of banana-induced rage, my parole officer introduced me to this kitchen marvel and my life was changed. No longer consumed by seething anger and animosity towards thick-skinned yellow fruit, I was able to concentrate on my love of theatre and am writing a musical play about two lovers from rival gangs that just try to make it in the world. I think I'll call it South Side Story.

Banana slicer...thanks to you, I see greatness on the horizon.

475 Comments | Was this review helpful to you?   Yes   No   Report abuse

32,689 of 33,741 people found the following review helpful

⭐⭐⭐⭐⭐ **Saved my marriage**
By Mrs Toledo on July 30, 2012
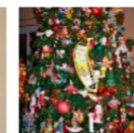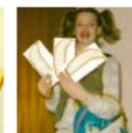
Size: 10" | Item Package Quantity: 1

What can I say about the 571B Banana Slicer that hasn't already been said about the wheel, penicillin, or the iPhone.... this is one of the greatest inventions of all time. My husband and I would argue constantly over who had to cut the day's banana slices. It's one of those chores NO ONE wants to do! You know, the old "I spent the entire day rearing OUR children, maybe YOU can pitch in a little and cut these bananas?" and of course, "You think I have the energy to slave over your damn bananas? I worked a 12 hour shift just to come home to THIS?!" These are the things that can destroy an entire relationship. It got to the point where our children could sense the tension. The minute I heard our 6-year-old girl in her bedroom, re-enacting our daily banana fight with her Barbie dolls, I knew we had to make a change. That's when I found the 571B Banana Slicer. Our marriage has never been healthier, AND we've even incorporated it into our lovemaking. THANKS 571B BANANA SLICER!

110 Comments | Was this review helpful to you?   Yes   No   Report abuse

13,193 of 13,761 people found the following review helpful

Customer Images

See all customer images

Most Recent Customer Reviews

⭐⭐⭐⭐⭐ **IT'S TOO YELLOW**
Love this product, however, I'm refrained of using it on the field because it is too bright. Enemy snipers can detect it from far away and almost got hit at least twice while...   Read more
Published 4 hours ago by pedro martin

⭐⭐⭐⭐☆ **Where do I put the toothpaste?**
I'm confused.
Published 17 hours ago by Amazon Customer

⭐⭐⭐⭐⭐ **Five Stars**
BEST white elephant gift ever! We gave it away with four bunches of bananas and everyone loved it.
Published 1 day ago by Stephanie

⭐⭐⭐⭐⭐ **Hutzler could it be that your reviews are more entertaining ...**
Hutzler could it be that your reviews are more entertaining that that puddle in Newcastle earlier today!? Keep this up and Netflix will have to up their game...
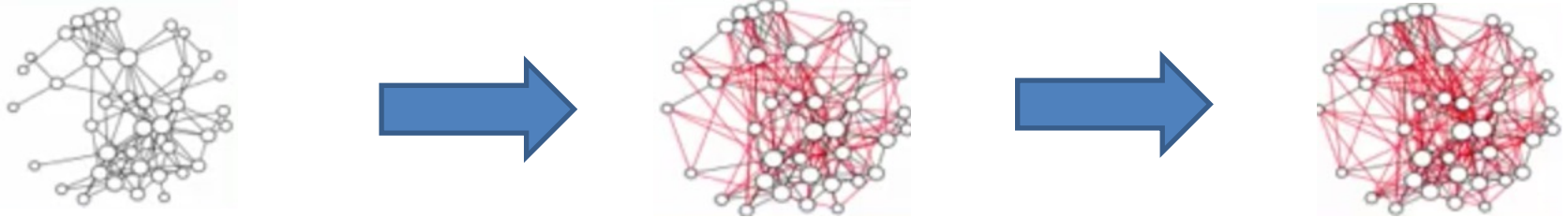Published 2 days ago by Amazon Customer

15

Valence = Measure of Connectivity

Valence increases over time

Presents social behavior and movements

- Challenges
  - More complex data exploration algorithms
  - Modeling and prediction of valence changes
  - Group event detection
  - Emergent behavior

Getting Value from Big Data

Descriptive Analytics   **Predictive Analytics**   **Prescriptive Analytics**

Uses data aggregation and data mining to provide insight into the past and answer:
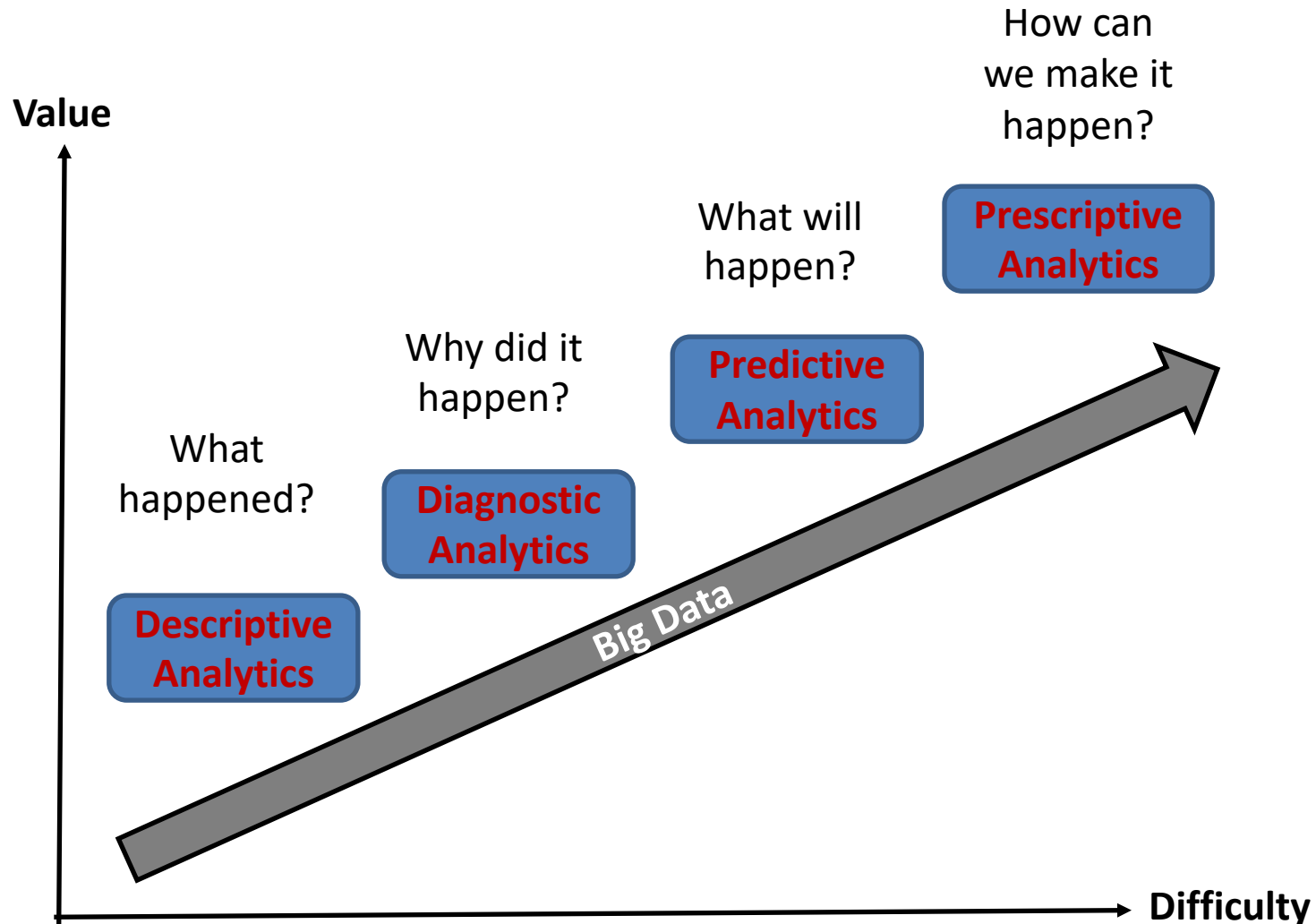

"What has happened?"

Uses statistical models and forecasts techniques to understand the future and answer:

"What could happen?"

Uses optimization and simulation algorithms to advice on possible outcomes and answer:

"What should we do?"

Getting Value from Big Data

Value

How can we make it happen?

Prescriptive Analytics

What will happen?

Predictive Analytics

Why did it happen?

Diagnostic Analytics

What happened?
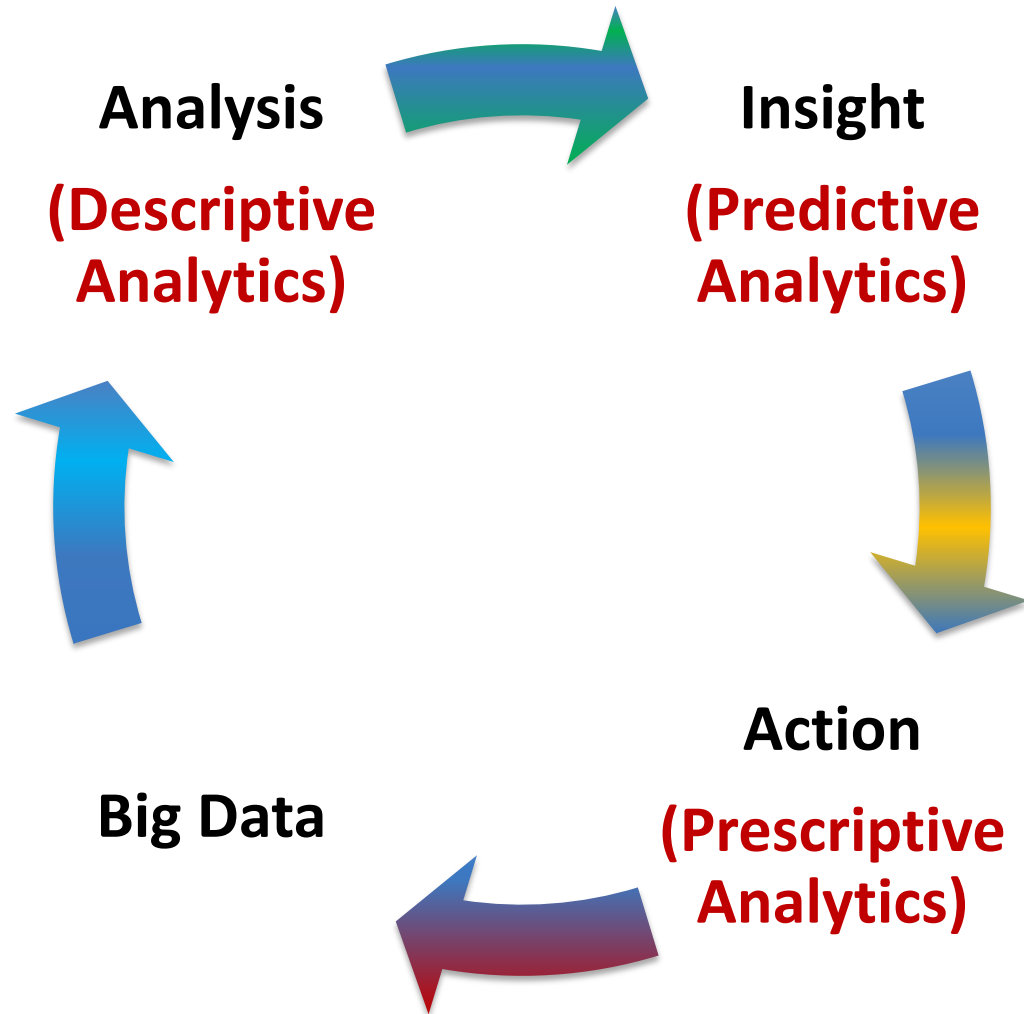
Descriptive Analytics

Big Data

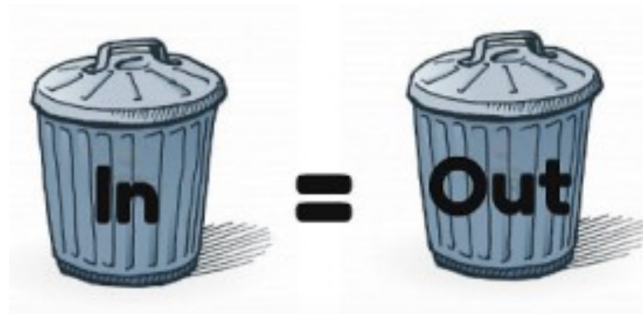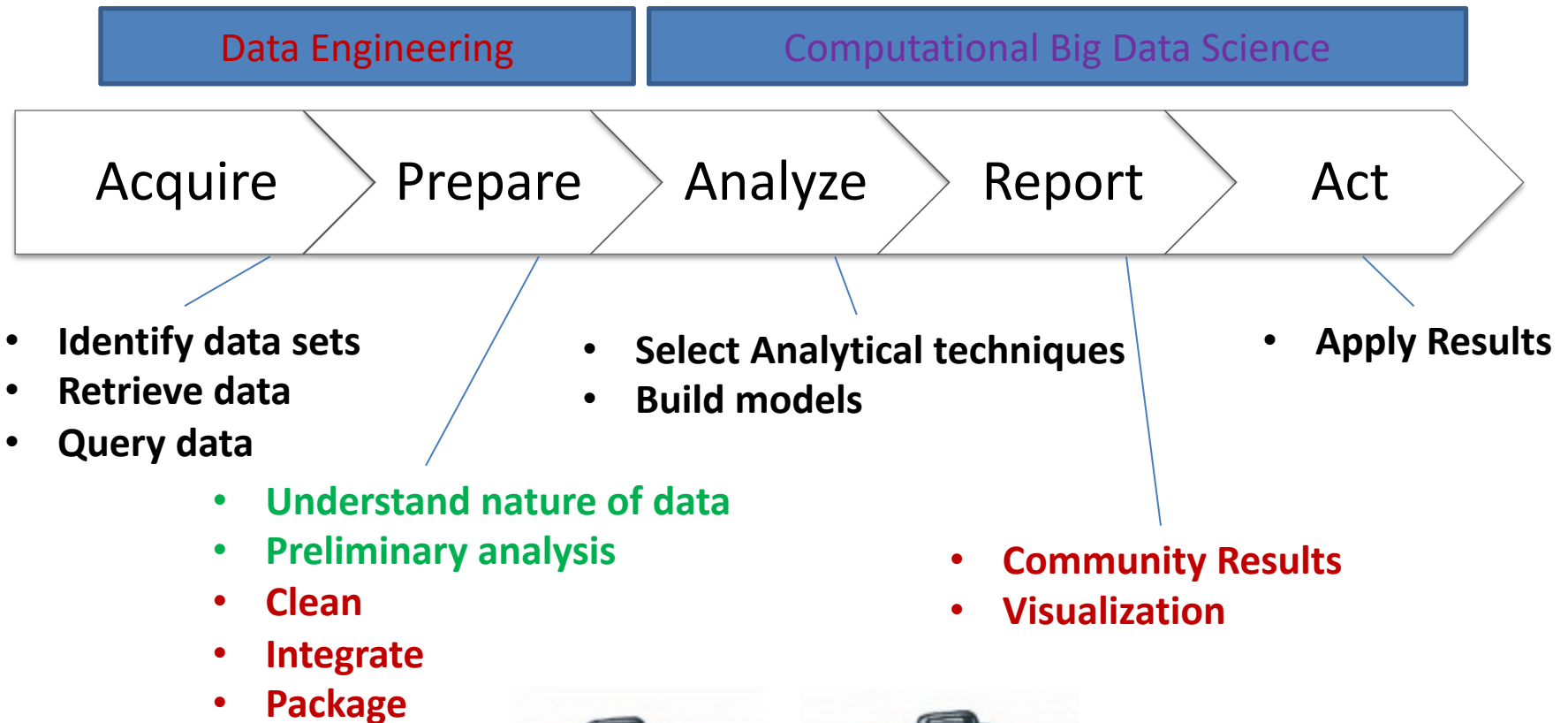**Difficulty**

Getting Value for Big Data

Descriptive Analytics: Insight into the past

Predictive Analytics: Understanding the future

Prescriptive Analytics: Advise on possible outcomes

Analysis
(Descriptive Analytics)

Insight
(Predictive Analytics)

Action
(Prescriptive Analytics)

Big Data

The Process of Data Analytics

Data Engineering	Computational Big Data Science

Acquire > Prepare > Analyze > Report > Act

- **Identify data sets**
- **Retrieve data**
- **Query data**

- **Select Analytical techniques**
- **Build models**

- **Apply Results**

- **Understand nature of data**
- **Preliminary analysis**
- **Clean**
- **Integrate**
- **Package**

- **Community Results**
- **Visualization**

20

Big Data Characteristics and Challenges

Questions?